LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-557152

A Bayesian Measurment Error Model for Misaligned Radiographic Data

K. P. Lennox, L. G. Glascoe

May 10, 2012

Technometrics

Disclaimer

A Bayesian Measurement Error Model for Misaligned Radiographic Data

Kristin P. Lennox,
and
Lee G. Glascoe *
Computational Engineering Division, Lawrence Livermore National Laboratory

Abstract

An understanding of the inherent variability in micro-computed tomography (micro-CT) data is essential to tasks such as statistical process control and the validation of radiographic simulation tools. These data present unique challenges to variability analysis due to the relatively low resolution of radiographs, and also due to minor variations from run to run which can result in misalignment or magnification changes between repeated measurements of a sample. Such positioning changes artificially inflate the variability of the data in ways that mask true physical phenomena. We present a novel Bayesian nonparametric regression model that incorporates both additive and multiplicative measurement error in addition to heteroscedasticity to address this problem. We use this model to assess the effects of sample thickness and sample position on measurement variability for an aluminum specimen. Supplementary materials for this paper are available online.

Keywords: Bayesian nonparametrics, Berkson error, curve registration, heteroscedasticity, micro-computed tomography, p-splines

*This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The experimental data collection and analysis were performed under sponsorship of the U.S. Department of Homeland Security, Science and Technology Directorate. The authors also wish to thank the Editor, Associate Editor and referees for their comments and suggestions.

1 Introduction

Micro-computed tomography (micro-CT) is a form of radiography employed for the nondestructive three-dimensional characterization of small objects. As in traditional CT scanning, multiple radiographs of an object are taken from different angles, and reconstruction algorithms are used to determine the properties of the three-dimensional sample. Micro-CT is applied in fields including industrial nondestructive testing (Hanke et al., 2008), forensics (Thali et al., 2003), paleobiology (Rossi et al., 2004), and biomedical research (Ritman, 2004).

An ongoing area of research is the simulation of radiographic systems. Successful simulation allows characterization of materials for which collecting experimental data is cost-prohibitive or otherwise infeasible, as well as the characterization of the measurement systems themselves for design purposes. The validation process for such models requires an understanding of the inherent variability of radiographic measurements so that comparisons between simulated and observed radiographs account for random variation in experimental results. Such variability assessments can also be used as part of a statistical process control scheme, to ensure that a machine is performing within acceptable tolerances. Both of these tasks require an understanding of how various sample and system characteristics influence measurement variability.

In order to achieve this understanding, repeated measurements of various samples taken over an extended period of time must be analyzed and the measurement variability quantified. Such a data set has been collected for the purposes of validating a computer model of a micro-CT machine; however, these data present some unique statistical challenges in the form of features such as positional measurement error and heteroscedasticity. Failure to account for these features would lead to incorrect conclusions about how measurement variability behaves in the micro-CT system. We have developed a novel Bayesian penalized spline (p-spline) model incorporating both measurement error and heteroscedasticity, which we have successfully applied to this problem. P-spline models employ a penalty term on the spline coefficients, rather than knot selection or derivative based penalties, to control the smoothness of the resulting function estimate (see e.g. Eilers and Marx, 1996; Ruppert and Carroll, 1999; Ruppert et al., 2003; Lang and Brezger, 2004). Their flexibility and rela-

tively simple implementation have led to their being employed for increasingly sophisticated modeling in both frequentist and Bayesian contexts, including measurement error problems (Carroll et al., 1999; Berry et al., 2002; Ganguli et al., 2005; Carroll et al., 2006) and situations requiring adaptive penalty terms (Ruppert and Carroll, 2000; Lang and Brezger, 2004; Jullion and Lambert, 2007; Krivobokova et al., 2008). Ruppert et al. (2003) give an iterative frequentist approach to incorporating heteroscedasticity into p-spline models, and Crainiceanu et al. (2007) demonstrated a Bayesian method. While a spline model incorporating heteroscedasticity and measurement error has been employed for density estimation (Staudenmayer et al., 2008), to the best of our knowledge this is the first model combining these elements in the regression context.

In addition to assessing the variability of the radiographs at various locations, we were also able to use the posterior distributions of the measurement error variables to discover additional features of the data. In particular, a magnification effect was identified that corresponded to a multiplicative measurement error situation. This in turn led to an enhanced measurement error model including both additive and multiplicative terms.

In Section 2 we describe the experimental setup of the relevant micro-CT system, the potential benefits to quantifying measurement variability for the system, and the difficulties in analysis caused by alignment error between samples. Section 3 contains a novel penalized spline model incorporating both heteroscedasticity in the response and additive measurement error in the explanatory variable that can be used to address the alignment error problem. In Section 4 we present the results of applying this model to the micro-CT data set, and also show an enhancement that incorporates multiplicative measurement error to account for a magnification effect. Section 5 contains a discussion of our findings.

2 Variability in Micro-CT Data

2.1 Description of the Micro-CT System

Figure 1 shows a side-view of the micro-CT machine under consideration. For this particular setup, x-rays emitted from the source pass through a 2-slit collimator. The top slit allows x-rays to pass through a test sample, while the lower slit gives access to a set of six reference

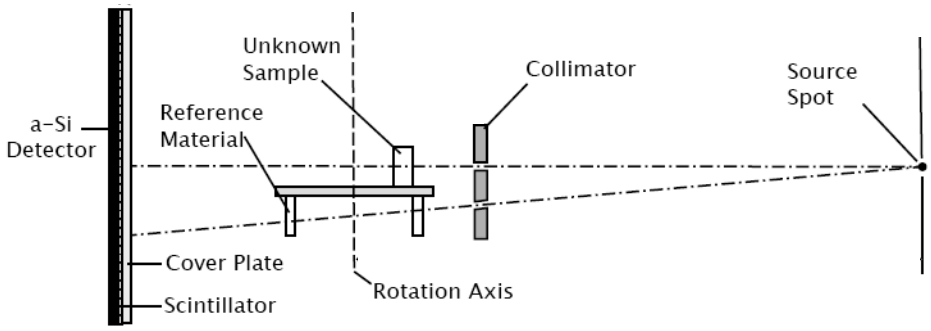

Figure 1: Diagram of an experimental configuration for a micro-CT machine.

specimens. These reference specimens are included in every micro-CT run for quality control purposes. Unabsorbed x-rays are registered by an amorphous silicon detector. Samples and reference specimens are attached to a carousel, which rotates in increments of half a degree to give 400 different views per experimental run.

Figure 2 shows different summary data from a single radiograph. Figure 2(a) is the measured transmittance at every pixel of the detector. Transmittance is I/I_0 where I is the measured intensity at a particular pixel and I_0 is the background intensity value corresponding to measured intensity when there is no intervening material. Transmittance is

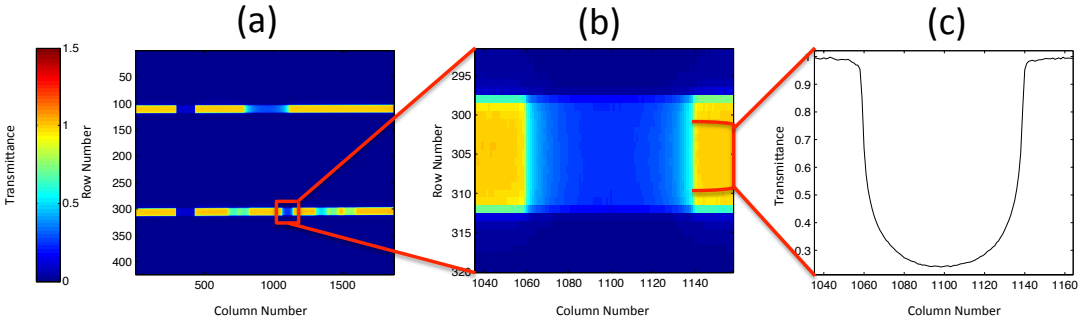

Figure 2: The process of converting a radiograph into a lineout for the aluminum reference specimen. These images show transmittance measurements, which are I/I_0. The remainder of this paper focuses on absorbance measurements, which are $-\ln(I/I_0)$. The online version of this figure is in full color.

close to 1 in regions where there is no intervening material between the source and the detector, and drops to 0 when the intervening material blocks all x-ray transmission. The two light horizontal bands correspond to the slits in the collimator between the source and the carousel. The darker vertical bands are caused by either the specimen of interest (top slit) or reference specimens (lower slit). The extremely dark vertical band on the far left is a copper strip attached to the collimator for calibration purposes. Figure 2(b) shows the isolation of a single reference specimen, in this case an aluminum cylinder, from the larger radiograph. Figure 2(c) is a "lineout": a one-dimensional summary of a radiograph that is in this case generated by the pointwise median from the central rows of the lower slit. Lineouts from multiple radiographs taken for different carousel positions can be used to generate 2-dimensional reconstructions, which are representations of a "slice" taken out of the core of the sample.

2.2 Benefits of Understanding Micro-CT Variability

An understanding of measurement variability is essential for any inferential procedure to ensure that a genuine result of interest can be distinguished from random chance. For the micro-CT system described, there are two particularly relevant applications for a comprehensive variability study of the system.

The first application is for the validation of the HADES radiographic simulation tool. A computational model for a radiographic system is useful for a variety of purposes, including supplementing physical measurements when large numbers of samples need to be characterized and exploring the potential effects of system changes on measurement quality. HADES is a ray-tracing radiographic simulation code developed under the auspices of the Department of Energy (Aufderheide et al., 2004). It uses pre-computed results from high-fidelity Monte Carlo simulators such as MCNP (Brown et al., 2002) to model x-ray source and detector behavior combined with a geometric model and the LLNL Evaluated Photon Data Library (Cullen et al., 1989) to account for the effects of x-ray passage through target objects. This strategy allows HADES to be used to rapidly simulate fairly large experimental setups with high accuracy. In the past, HADES has been applied to problems such as assessing the relative impact of different sources of radiographic blur (von Wittenau et al., 2002). There is currently interest in using HADES to supplement experimental measurements for the partic-

ular micro-CT system described in this paper, which would lead to significant cost reductions as opposed to carrying out physical measurements for large numbers of samples (Chen et al., 2011). Before HADES can be used in this way, the accuracy of the simulator must be quantified for a variety of materials. However, the existing data intended for validation consists of materials that were each characterized by only a single micro-CT run. Due to changes in the experimental setup it is not possible to acquire additional runs for these validation specimens, and therefore it is not possible to quantify the measurement variability of the machine using only the original validation data. Without an understanding of measurement variability, discrepancies between a HADES simulation and the corresponding measurement can not be definitively attributed to errors in the simulation, since they could be due to fluctuations in the micro-CT system. Some method for estimating measurement variability would allow HADES to be evaluated more fairly.

A second application is for online quality control for the micro-CT machine itself. An automated method could compare single lineouts for the existing quality control specimens to a distribution based on a sample of preexisting measurements. If the system produces results which are inconsistent with previous measurements, this is an indication that there could be a problem with the machine. At that point, diagnostics could be run to identify why the system is producing inconsistent results. In addition to flagging suspect results in an automated manner, this method could save time by providing the means to identify aberrant runs before they are complete.

Both of these applications require a solid understanding of the behavior of micro-CT measurements produced by a particular machine over the course of many experimental cycles taking place over an extended period of time. This, in turn, requires repeated measurements for a set of specimens. Fortunately, the micro-CT system has such a dataset in the form of the reference specimens from the lower level of the carousel. Each reference specimen was measured repeatedly over the course of numerous micro-CT runs as a quality control measure, and the presence of multiple materials in the reference specimen set provides an opportunity to study the overall behavior of the system. These specimens can also form the basis for an online quality control scheme once a variability study is completed.

Rather than presenting the results of the full analysis of the reference specimen set, this paper instead focuses on the statistical challenges inherent in that analysis. For this reason,

the remainder of the paper will focus on the aluminum reference specimen. Note that the challenges identified in the analysis of this single specimen are characteristic of those seen for the larger set, and, indeed, of any repeated measurements analysis for this system. Therefore, the methods presented here are suitable for, and indeed intended for, wider application.

2.3 Description of Aluminum Reference Specimen Data

The data under consideration consist of 28 micro-CT runs of an aluminum cylinder. The aluminum reference specimen is of particular interest in that it frequently passed through both the location closest to the source and the location closest to the detector in a single micro-CT run. Since runs cover only 200° of rotation, this behavior was not observed for any other sample measured multiple times. Thus, studying the aluminum sample allows us to assess unique positional effects on measurement variability.

Three radiographs from each run were selected for assessment, corresponding to the 0°, 90°, and 180° carousel positions, where 0° is the location placing the sample closest to the x-ray source, and 180° is the location closest to the detector. For this analysis, transmittance measurements are converted to absorbance, which is $-\ln(\text{transmittance})$ or $-\ln(I/I_0)$. These absorbance radiographs were then converted to median filtered lineouts, which consisted of the pointwise median of five pixels from the core of the specimen for each column in the radiograph. Lineouts for this application were 161 detector pixels long, which encompassed the sample of interest as well as some background areas in all cases. Lineouts were considered in place of full radiographs due to the fact that the median filtering reduced variability, which should provide more informative results for either comparisons to HADES or for quality control purposes. Median filtering is preferred to mean filtering when, as in this application, an image may contain extreme outliers in the form of bad detector pixels.

The pixel-wise mean and standard deviation for each of the three carousel positions are shown in Figure 3. The mean lineouts for the aluminum sample are similar between the 3 locations. However, as the sample moves farther from the source, the portion of the lineout containing the sample becomes smaller. This is due to the fact that samples closer to the source have a greater effective magnification than those farther away, and also possibly due to the increased effect of x-ray scatter. Also of interest are the standard deviation spikes at the edges of the sample that appear for each carousel position. A naïve interpretation

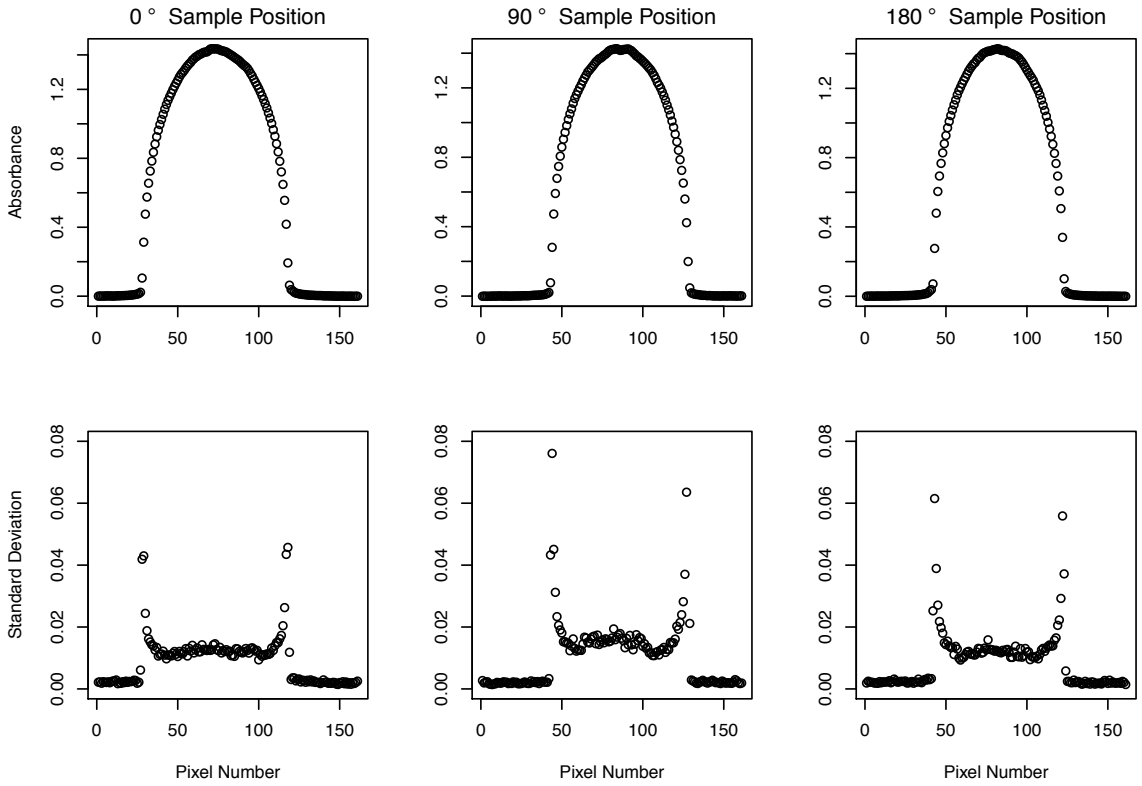

Figure 3: Mean and standard deviation at each pixel for an aluminum sample measured at three different locations relative to the x-ray source and detector.

of these results would be that variability is exceptionally high for very thin portions of a sample, but stabilizes as sample thickness increases. However, the spikes are smaller for the sample with the greatest magnification (0° position), suggesting that this may be an artifact of relatively low lineout resolution.

Further investigation reveals this to indeed be the case. Figure 4 shows an overlay of two lineouts from the 180° carousel position, along with the pointwise difference between them. The first sample is shifted slightly to the left of the second sample, as can be seen particularly clearly at the sample edges. On the left edge, the grey measurement from lineout 1 is lower than the aligned black measurement from lineout 2, while on the right side of the sample the relative sizes of the measurements are reversed. However changing the alignment by a full pixel would result in a worse overall alignment. Such shifts are due to small changes in the starting configuration in the micro-CT system from run to run. These are corrected for the aluminum data using a least squares alignment, but this alignment cannot correct for

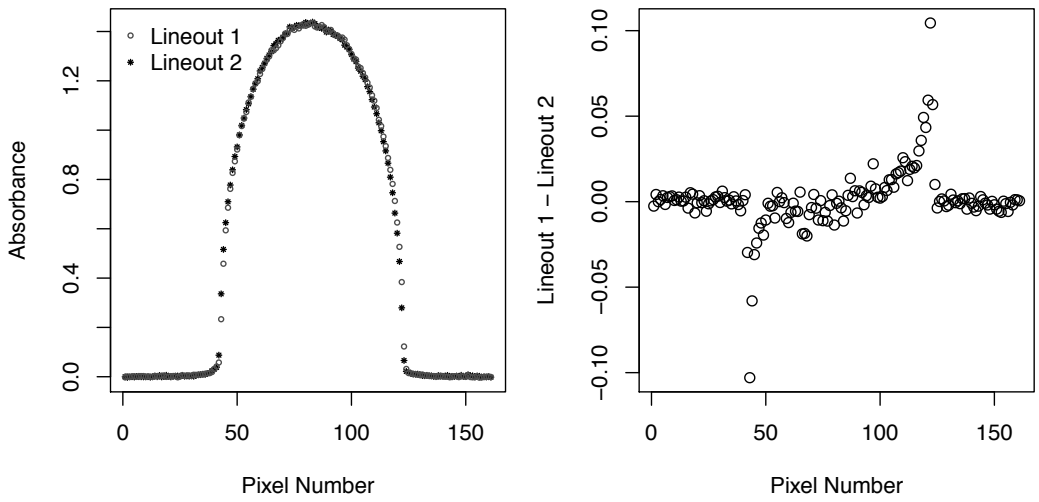

Figure 4: Two lineouts from the 180° carousel position showing evidence of misalignment. The second plot shows the pointwise difference between these two lineouts.

residual error on the order of a half pixel or less. This residual misalignment has little effect in the background and core of the sample, where the derivative of the absorbance curve is near 0, but will contribute significantly to variability at the edges of the sample. Since a comprehensive variability analysis requires an understanding of how absorbance variability is affected by sample thickness, it is necessary to separate out this misalignment effect as much as possible.

The misalignment problem may be viewed as an example of Berkson error (Berkson, 1950). For an observed surrogate P, true covariate value X, and error term W, Berkson error satisfies the following relationship:

$$X = P + W. \tag{1}$$

In this case the observed value is the pixel location, while the true covariate is the actual location relative to the sample for which absorbance was measured. In contrast to classical measurement error, in the Berkson error situation, the surrogate value is less variable than the value of the true covariate. This is clearly true in this case, where the observed pixel number can be considered to be a "rounded" version of the true location with respect to the sample. Berkson error is known to substantially alter residual variance estimates in some cases, even when the mean estimate is largely unaffected (Carroll et al., 2006). Thus, given

our interest in the variance for this particular problem, the presence of Berkson error is a cause for concern.

3 Bayesian Spline Model with Measurement Error

The model employed to analyze these data must have several key properties. The model for the mean must capture all relevant features for the radiographic lineouts. The model must allow for heteroscedasticity from pixel to pixel to capture the changing characteristics of variability throughout a sample. For this application the model for the variance is as important, if not more so, than the model for the mean. In such cases it is imperative that a model which can adequately capture the behavior of the variance be used (Carroll, 2003). Finally, the model must incorporate some method for correcting the misalignment discussed in the previous section.

We selected a Bayesian penalized spline (p-spline) approach for our base model. The advantages of using the Bayesian context for the model are two-fold: 1) Bayesian p-splines have already been used for both measurement error models and in situations where there is heteroscedasticity in the response error and provide a natural way to integrate both, and 2) Bayesian models provide uncertainty information about all variables of interest, including the local variance and measurement error, in the form of the posterior distribution.

Our model is based on that of Crainiceanu et al. (2007), which employed both adaptive penalty terms for a p-spline regression model and heteroscedasticity in the response, and we have used a modified form of their MATLAB code for this analysis. We have altered this model to incorporate both a Berkson error component and positional independence for heteroscedasticity. The new model uses a penalized linear spline with knots $\{\kappa_k^m\}_{k=1}^{K_m}$. It can

be summarized as:

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_{ij} + \sum_{k=1}^{K_m} b_k (x_{ij} - \kappa_k^m)_+ + \epsilon_{ij} \\
\beta_k &\sim \mathrm{N}(0, \sigma_\beta^2) \\
b_k &\sim \mathrm{N}(0, \sigma_b^2(\kappa_k^m)) \\
\epsilon_{ij} &\sim \mathrm{N}(0, \sigma_\epsilon^2(p_j)) \\
x_{ij} &= p_j + w_i \\
w_i &\sim \mathrm{Unif}(-0.5, 0.5)
\end{aligned}
$$

where y_{ij} is the measured absorbance for the ith lineout at an observed pixel location p_j, x_{ij} is the true location with respect to the specimen where the measurement was taken, and $(\cdot)_+ = \max(\cdot, 0)$. The measurement error w_i is shared by all pixels in a particular lineout, and is modeled as coming from a uniform distribution on the interval (-0.5,0.5). The magnitudes of the offset terms are restricted due to the lack of identifiabilty of the offsets ($\{w_i\}_{i=1}^n$ is equivalent to $\{w_i + \gamma\}_{i=1}^n$). By limiting the permissible range, offset "drift" in the Markov Chain Monte Carlo (MCMC) sampler is prevented. The 1 pixel range was found to be adequate to capture the relative magnitudes of all offsets for these data, which agrees with the intuition that the initial least squares alignment step should eliminate misalignment greater than half a pixel in either direction. For problems where hard boundaries for offset sizes may not be appropriate, alternative solutions to the identifiability problem include setting one offset to 0 or restricting the mean of the collection of offsets to be 0. Similar results were observed for all three alternative prior formulations when applied to micro-CT data.

Note that the effective smoothing parameter for a p-spline with constant error variance σ_ϵ^2 and constant σ_b^2 is the ratio $\sigma_b^2/\sigma_\epsilon^2$, with a small value corresponding to a large amount of smoothing. Therefore, if heteroscedastic errors with variance $\sigma_\epsilon^2(\cdot)$ are included in the model without an adaptive $\sigma_b^2(\cdot)$ term, then the effective smoothing parameter will vary according to the estimated error variance, but not the behavior of the underlying mean function. Allowing both terms to be functions of the input location provides more flexibility for roughness penalization.

The prior on $\sigma_b^2(\kappa_k^m)$ follows a degree 1 log spline model as in the original Crainiceanu

et al. (2007) paper:

$$\log(\sigma_b^2(\kappa_k^m)) = \gamma_0 + \gamma_1\kappa_k^m + \sum_{r=1}^{K_b} c_r(\kappa_k^m - \kappa_r^b)_+ + v_k$$

$$\gamma_r \sim \mathrm{N}(0, \sigma_\gamma^2)$$

$$c_r \sim \mathrm{N}(0, \sigma_c^2)$$

$$1/\sigma_c^2 \sim \mathrm{Gamma}(\alpha_c, \beta_c)$$

$$v_k \sim \mathrm{N}(0, \sigma_v^2).$$

The error precision $1/\sigma_\epsilon^2(p_j)$ at pixel j is treated independently of other pixels, and is given a gamma prior:

$$1/\sigma_\epsilon^2(j) \sim \mathrm{Gamma}(\alpha_\epsilon, \beta_\epsilon).$$

This is somewhat reminiscent of the method in Lang and Brezger (2004), where the smoothing parameter $\sigma_b^2(\cdot)$ was treated independently from knot to knot. The distinction lies in the fact that our model focuses on highly localized behavior in the error variance, while Lang and Brezger (2004) were concerned with localized behavior in the mean function.

Variance parameters at different pixel locations are treated independently due to concern that smoothing the variance estimates within the model could mask genuine characteristics of the data. In situations where variance would be expected to change smoothly with respect to the covariate, the original log spline model for error variance from Crainiceanu et al. (2007) would be appropriate.

Details of the MCMC procedure used for this model, including prior parameter specifications, sampling details, and convergence diagnostics, are found in the Supplementary Section S.1.1. For a MATLAB implementation, 3,000 iterations took roughly 40 minutes to run on a MacPro desktop with dual quad core 3.2 GHz processors. A similar model not incorporating measurement error ran in roughly 30 minutes on the same machine.

4 Application for Misaligned Micro-CT data

As discussed in Section 2, variability measurements for micro-CT lineouts are influenced by small amounts of alignment error. The penalized spline model described in Section 3

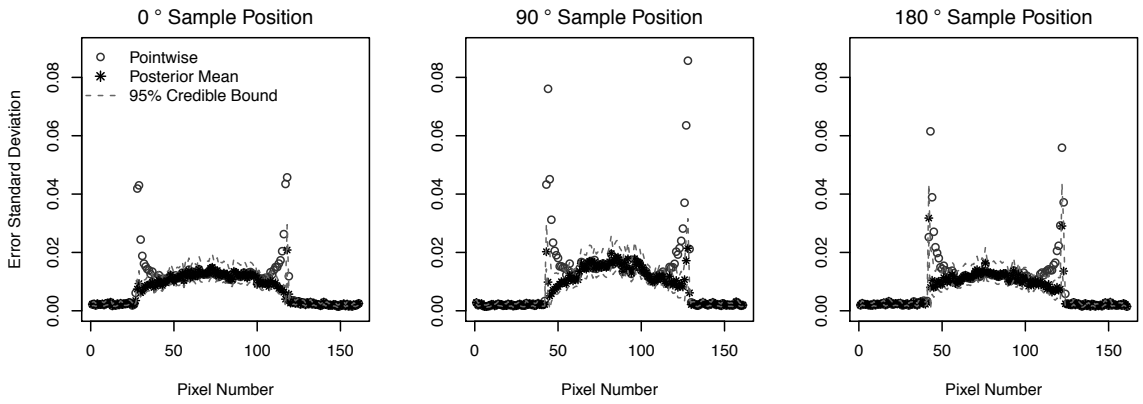

Figure 5: The posterior mean for the error standard deviation compared to raw pointwise results for all three data sets. Pointwise 95% credible intervals are shown as dashed lines.

provides a method for simultaneously correcting for alignment error while estimating pixel-wise variability for a micro-CT lineout.

4.1 Spline Model with Measurement Error

The Bayesian spline model including measurement error was fit to all three position data sets for the aluminum specimen. The estimates for the mean absorbance are very similar to the pointwise results, indicating that misalignment error does not have a substantial effect on the mean in this case. This is illustrated in Supplementary Figure 3.

The posterior mean for the standard deviation ($\sigma_\epsilon(p_j)$) for absorbance is shown in Figure 5. In this case, the difference between the measurement error model and raw results is more dramatic. The posterior standard deviation estimates at the edges of the sample are substantially less than those based on raw data, with the largest change being an 10-fold reduction for pixel 117 in the 0° sample. In locations where the slope of the absorbance function is near 0, we once again see agreement between the raw pointwise results and those corrected for alignment error. Some residual variability spikes remain at the edges of the sample. One suspected cause is further explored in Section 4.2. The most complete correction of the spiking phenomenon occurs for the 0° sample, which is perhaps unsurprising since it suffered the least from misalignment error in the first place.

4.2 Magnification Effects and Residual Variability

It has been noted that the variability spikes at the edges of the sample are significantly reduced by the alignment error model, but are not eliminated completely. Some portion of this variability was suspected to be due to magnification differences, which manifest as alternative alignments. The magnification of a specimen for a particular radiograph is a function of its distance from the detector. The closer a sample is to the detector, the smaller its magnification will be. This is why the lineouts for the 180° data set, which corresponds to the location closest to the detector, are narrower at the base than those in the 0° set, for which the sample was farthest from the detector. Even for a fixed carousel position, slight changes in specimen location, and therefore magnification, are observed.

Since two lineouts with different magnification cannot be perfectly aligned against each other, they tend to align preferentially against one side. This leads to a dramatic reduction in observed variability on the preferentially aligned side, but not on the other. (Note that magnification errors tend to be smaller than misalignment errors, so some reduction will be observed on both sides relative to the raw pointwise results.) Our model can capture such alternative alignments and include them in posterior estimates for the mean and variance. An example of alternative alignments can be seen in Figure 6, which shows an alternative alignment captured by one of the Markov chains for the 90° sample position. In the left plot, mean offsets for each position are shown. Due to a lack of identifiability, each collection of 28 offsets is standardized to have mean 0. Notice that three of the chains show highly similar alignments, while the fourth chain exhibits a distinctly different conformation that at times differs from the more common one by more than a tenth of a pixel. The corresponding effect on posterior standard deviation estimates is shown in the right plot of Figure 6. The standard chains preferentially correct for misalignment on the left side of the sample, leading to a reduction in variability relative to the alternative chain for most points between pixel numbers 40 and 80. Similarly, the alternative alignment leads to a reduction in standard deviation on the right side of the sample for most pixel locations between 80 and 120. All chains still show sharp increases in standard deviation within two pixels of the edge. Note that we have never observed a shift from one preferred alignment to another within a Markov chain. Due to the relatively slow mixing of the alignment offsets, the most efficient way we

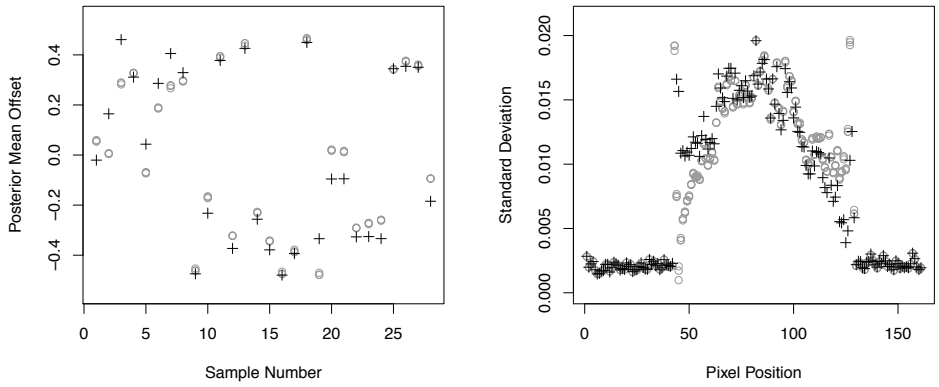

Figure 6: A demonstration of alternative alignments for the 90° sample. Results from three Markov chains showing one alignment are denoted by circles, while the Markov chain showing an alternative alignment is denoted by a '+'.

have found to identify these situations is by using randomly initialized offsets and running multiple chains.

This issue also illustrates the value of a full Bayesian analysis. A point estimate of measurement error, such as a maximum likelihood or maximum a posteriori estimate, would fail to capture this behavior, and lead to an artificial increase in the asymmetry of the results, as well as failing to present an explanation for this source of residual alignment error.

4.3 Spline Model with Both Additive and Multiplicative Measurement Error

To address the presumed magnification effect, a model that combined additive and multiplicative error was implemented. Specifically, the model from Section 3 was modified such that:

$$x_{ij} = m_i(p_j - c_i) + c_i + w_i$$
$$m_i \sim N(1, \sigma_m^2)$$

where m_i is a multiplicative measurement error term that represents a change in the magnification of the sample. The term c_i is a deterministic value for the center of the aluminum

sample in a particular lineout, and is identified by taking the median of all pixel locations with absorbance greater than half of the maximum absorbance for the lineout. This method was found to be more robust than taking the location of the global maximum for absorbance. The centering variables are not required to correct for multiplicative measurement error; however they do ensure that the additive terms w_i in this model are directly comparable to those from the version without multiplicative measurement errors. In their absence, the values for w_i become larger since the alignment of the samples is affected by the rescaling as well as physical position shifts. MCMC details are found in Supplementary Section S.1.2.

Figure 7 shows the error standard deviation estimates from this model as functions of both pixel location and pointwise mean absorbance. Pointwise results and the estimates from the offset only model are included for comparison. The mean absorbance results were not substantially different from either the pointwise or offset only results, and are not pictured.

The error standard deviation plots reveal several interesting features. Firstly, both models including measurement error components, represented in Figure 7 by black and grey stars, provide substantial reductions in variability as compared to the pointwise results, represented by grey circles. At the extreme edges of the sample, all three models show variability spikes, with the largest for the pointwise results, the second largest for the magnification model results, and the smallest for the offset only model. This result indicates that uncorrected magnification changes are not causing the residual variability spikes observed in the offset only model, and indeed it seems that the inclusion of a magnification effect is producing a reduction in model accuracy. Potential causes for these spikes are discussed in Section 5. Moving slightly closer to the central portion of the specimen, such as the region from pixel 40 to pixel 50 for the 0° sample position, shows that the model including magnification reduces variability in this region. Not only that, but the variability of the standard deviations themselves, as represented by 95% credible bounds, is reduced relative to the offset only model. For example, the dashed bounds of the offset only model are considerably wider than the corresponding solid bounds for the new model at pixels 45-60 of the 180° position plot. This is at least in part due to the elimination of alternative alignments, which were never observed in the output of the model including a magnification effect. If the spiking at the edges of the sample is ignored, both standard deviation estimates and the width of credible bounds for those estimates gradually increase from background regions to the core of the

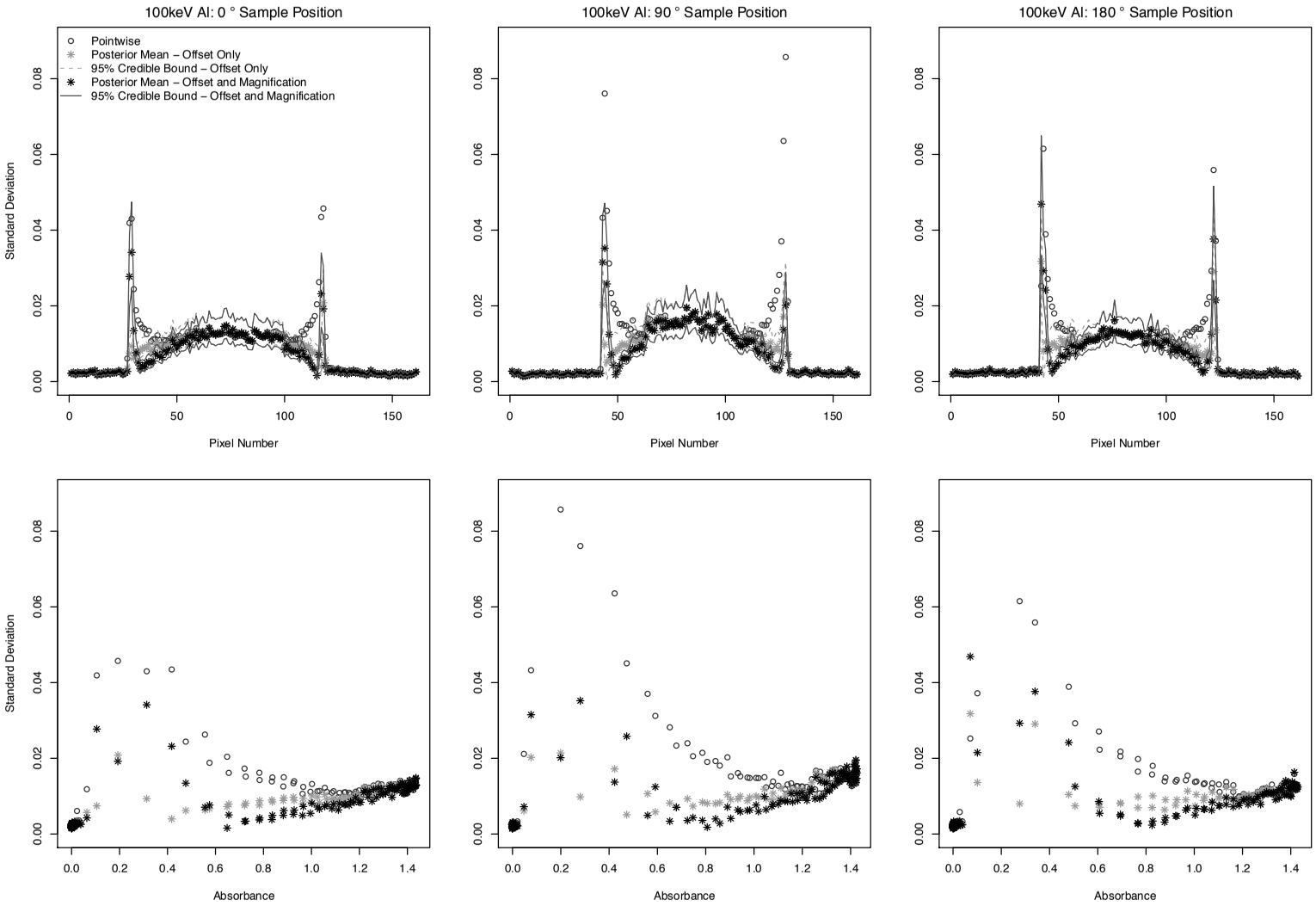

Figure 7: The posterior standard deviation results for models incorporating an offset or an offset combined with a magnification effect compared to raw pointwise results for all three data sets.

DIC Comparison

	0° Position	90° Position	180° Position
No Measurement Error	-31570	-30930	-33314
Offset Only	-33940	-33970	-34760
Offset and Magnification	-34407	-34462	-35342

Table 1: DIC comparison of models with and without measurement error components.

sample.

Accounting for alignment error, either multiplicative or additive, in the model radically changes the observed relationship between absorbance and variability. Note that absorbance in a uniform material has a monotonic relationship with the path length through the material, and to some extent can stand as a proxy for sample thickness.

The raw data results show a low level of background variability which spikes at the edges of the sample, and then decreases in a nonlinear fashion. It appears to be leveling off within the sample core. In contrast, for the results from both of the measurement error models, the standard deviation is increasing from the background rate to a higher level in the core, excluding residual variability spikes for 3-5 pixels on either edge of the sample. This leads to the fundamentally different conclusion that variability in absorbance measurements will generally increase as sample thickness increases, and once again highlights the importance of addressing measurement error when estimating residual variances.

It is not immediately obvious which measurement error model, additive only or additive and multiplicative, provides the best fit to the data. While the model including the magnification effect produces additional variability reduction in the core of the sample, it increases observed variability at the extreme edges. Therefore, formal model selection procedures are needed to evaluate the relative quality of these models. It was decided to also include a spline model without any measurement error terms in the comparison. (Additional details on the no measurement error model are given in Supplementary Section S.2.) The Deviance Information Criterion (DIC) proposed by Spiegelhalter et al. (2002) provides a natural means of comparing multiple Bayesian models fit via MCMC, where the smallest DIC value corresponds to the best model. More details on the DIC are provided in Supplementary Section S.4. The results of this analysis are given in Table 1.

For all three specimen positions, the model with magnification error is superior to the other models, as indicated by its having the lowest DIC. Similarly in all three cases, the offset only model is superior to the no measurement error model. This provides compelling evidence that the inclusion of both offset and magnification effects in the model for these data is justified.

4.4 Position Effect

Recall that the particular value of the aluminum reference specimen is that it is the only specimen with repeated measurements available at both the carousel position closest to the source and that closest to the detector. This allows us to evaluate what difference, if any, the location of the sample makes on measurement variability. Figure 8 shows a comparison of the variability results (posterior mean variance) for each of the three sample positions and for both measurement error models. Note that variance is pictured instead of standard deviation due to the fact that it shows a better separation of both model and positional results.

High variance values in the 0.1-0.5 absorbance region correspond to remaining variabil-

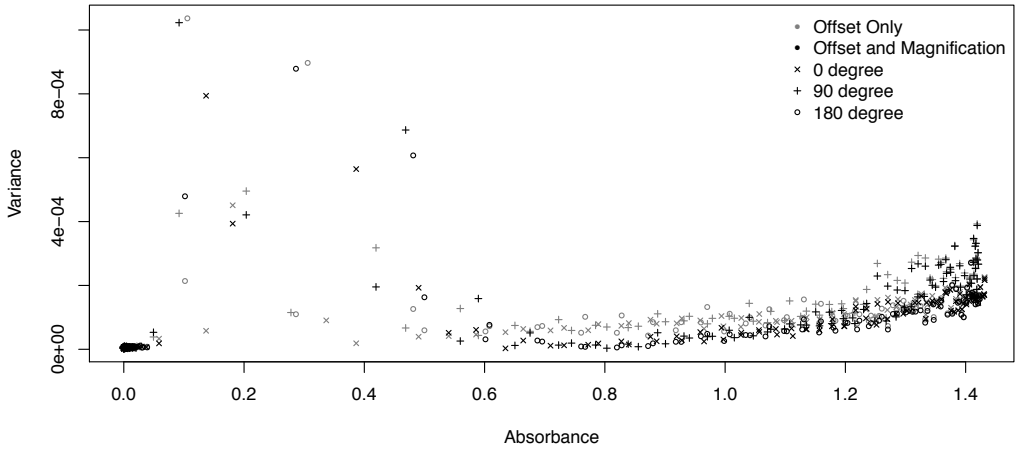

Figure 8: Comparison of measurement variability for each of the three sample positions. The plot includes both the final model which corrects for both magnification and offset errors as well as the offset only model.

ity spikes at sample edges, which will be further discussed in Section 5. Interestingly, all specimen positions have similar error variance values to one another throughout most of the absorbance range regardless of the measurement error model. The results for the offset only and offset plus magnification models are clearly separated throughout most of the range. For both models, however, the 90° position data, indicated by the '+' character in the figure, has a higher variance in the sample core (i.e. the highest absorbance region) than either the 0° or 180° samples. This could be due to the fact that the 90° sample position corresponds to a different location on the detector, which might have higher measurement variability. In any case, samples measured close to the center of the detector appear to have lower measurement variability than those at the edge, regardless of the exact distance from the detector and source.

5 Discussion and Conclusions

We have presented a new model for nonparametric regression in the presence of both heteroscedasticity and measurement error. This model was developed for and applied to a radiography data set that exhibited misalignment that severely affected variability estimates at the edges of samples. An additional source of misalignment, in the form of a magnification effect caused by the sample moving closer to or farther from the x-ray source was also identified, and addressed via a multiplicative measurement error term. When misalignment was removed, variability in the absorbance measurements was reduced significantly, revealing an apparent monotone increasing relationship between absorbance (which is a proxy for sample thickness) and the standard deviation of absorbance measurements. Some residual variability spiking was still observed at the extreme edges of specimens regardless of the alignment technique applied.

By comparing results for the same specimen at three carousel positions, it was also discovered that measurement variability was not uniform across all locations. In particular, measurements taken at the center of the detector (0° and 180° positions) exhibited lower core variability than measurements for the same specimen taken at the edge of the detector (90° position). This has implications for both quality control and simulation validation efforts, since it indicates that certain regions of the detector produce more repeatable radiographic

measurements than others.

The statistical techniques developed for this analysis are being applied to a larger reference specimen data set in an effort to develop a more comprehensive understanding of measurement variability for the micro-CT system. This variability analysis will ultimately be of use for applications such as the validation of radiographic simulations and the implementation of quality control procedures for the system.

Note that what we describe as measurement error modeling is equivalent to curve registration procedures in functional data analysis (see e.g. Ch. 7 Ramsay and Silverman, 2005). In particular, the additive measurement error component corresponds to a fully Bayesian shift registration model, and the combined multiplicative and additive measurement error model is a linear registration transformation. While more general warping functions, Bayesian and otherwise, appear in the literature (see e.g. Kneip and Ramsay, 2008; Telesca and Inoue, 2008; Claeskens et al., 2010) and could perhaps reduce the residual spiking left after the application of either of our models, an advantage of the linear method we use lies in the physical interpretations of the additive and multiplicative terms. In particular, the additive component corresponds to jitter perpendicular to the line between the x-ray source and detector, while the multiplicative effect indicates movement parallel to this line. For the dual purposes we propose for our variability study, identifying these effects is of greater interest than a more general, but less interpretable, registration function. In particular, for HADES validation horizontal jitter is irrelevant, since HADES simulations can be carried out at arbitrarily high resolution. Understanding the range of magnification effects observed for reference specimens can help to highlight where lack of correspondence between a HADES prediction and a physical measurement is due to routine changes in sample location versus a potential significant error in model geometry. However, the the accuracy of the physical interpretation of the multiplicative error as a pure magnification effect is somewhat in doubt.

The wider functional data literature gives us cause for caution in the interpretation of the multiplicative error term we use to represent a magnification effect. It is well established that registration functions can produce unrealistic results due to the confounding of scaling differences with amplitude differences (Ramsay and Li, 2002). While the alternative alignment results from Section 4.2 provide strong evidence that some kind of magnification effect is in play, it is possible that the effect is smaller than that identified by the measure-

ment error model, with the excess representing corrections in amplitude differences rather than in the location of the sample itself. Evidence for this hypothesis is provided by the fact that the correlation between maximum absorbance for a particular observation with the corresponding posterior mean multiplicative error parameter is around -0.75 for all specimen locations. This could explain the increase in variability spiking observed in the magnification model as compared to the offset only model. Correcting for amplitude changes closer to the core of the sample, and thereby reducing variability in that region, could be reducing the quality of the fit for extreme edge pixels in an un-physical way. It also indicates that the geometric interpretation of the multiplicative measurement error variables should be used with caution, since the physical magnification component may differ from the corresponding multiplicative factor. A possible solution to this problem would be to study the physical micro-CT system to determine what range of magnification effects is physically reasonable, and impose prior constraints to limit the measurement error model to that range. This would correspond to the curve registration technique of imposing strong regularity conditions on warping functions. We also continue to explore other possible physical causes for remaining misalignment and methods to incorporate them into future modeling.

The final point we wish to make relates to the importance of understanding the character of data when carrying out statistical analyses. In the absence of alignment correction, the pointwise variability results give a misleading impression of the relationship between sample thickness and the variance of the measured absorbance. If the apparent asymptote in variability in the core of the sample had been taken as real rather than an artifact, then this would have caused a fundamental misunderstanding of the behavior of absorbance measurements, and could potentially have lead to incorrect inference for new samples of different diameters. By determining that a measurement error situation was causing an artificial increase in measurement variability and correcting for it, a materially different answer was identified. This example emphasizes the need for a careful understanding of data prior to selecting and applying inferential procedures.

SUPPLEMENTARY MATERIAL

MCMC and Model Selection Details: Supplementary figures for model fits and convergence diagnostics, as well as additional details on MCMC implementation and model

selection. (PDF)

Software: MATLAB code will be posted online when available. (ZIP)

References

Aufderheide, M., Henderson, G., von Wittenau, A., Slone, D., and Martz Jr, H. (2004), "HADES, a code for simulating a variety of radiographic techniques," in *Nuclear Science Symposium Conference Record*, IEEE, vol. 4, pp. 2579–2583.

Berkson, J. (1950), "Are there two regressions?" *Journal of the American Statistical Association*, 45, 164–180.

Berry, S., Carroll, R., and Ruppert, D. (2002), "Bayesian smoothing and regression splines for measurement error problems," *Journal of the American Statistical Association*, 97, 160–169.

Brown, F., Barrett, R., Booth, T., Bull, J., Cox, L., Forster, R., Goorley, T., Mosteller, R., Post, S., Prael, R., et al. (2002), "MCNP version 5," *Trans. Am. Nucl. Soc*, 87, 4.

Carroll, R. (2003), "Variances are not always nuisance parameters," *Biometrics*, 59, 211–220.

Carroll, R., Maca, J., and Ruppert, D. (1999), "Nonparametric regression in the presence of measurement error," *Biometrika*, 86, 541–554.

Carroll, R., Ruppert, D., Stefanski, L., and Crainiceanu, C. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*, vol. 105, CRC Press.

Chen, H., Aufderheide, M., White, W., Roberson, G., and Glascoe, L. (2011), "Validation of HADES-based simulations of radiographic experiments," in *2011 IEEE International Conference onTechnologies for Homeland Security*, IEEE, pp. 104–109.

Claeskens, G., Silverman, B., and Slaets, L. (2010), "A multiresolution approach to time warping achieved by a Bayesian prior–posterior transfer fitting strategy," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 673–694.

Crainiceanu, C., Ruppert, D., Carroll, R., Joshi, A., and Goodner, B. (2007), "Spatially adaptive Bayesian penalized splines with heteroscedastic errors," *Journal of Computational and Graphical Statistics*, 16, 265–288.

Cullen, D., Chen, M., Hubbell, J., Perkins, S., Plechaty, E., Rathkopf, J., and Scofield, J. (1989), "Tables and graphs of photon-interaction cross sections from 10 eV to 100 GeV derived from the LLNL evaluated photon data library (EPDL)," Tech. rep., Lawrence Livermore National Lab., CA (USA).

Eilers, P. and Marx, B. (1996), "Flexible smoothing with B-splines and penalties," *Statistical Science*, 89–102.

Ganguli, B., Staudenmayer, J., and Wand, M. (2005), "Additive models with predictors subject to measurement error," *Australian & New Zealand Journal of Statistics*, 47, 193–202.

Hanke, R., Fuchs, T., and Uhlmann, N. (2008), "X-ray based methods for non-destructive testing and material characterization," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 591, 14–18.

Jullion, A. and Lambert, P. (2007), "Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models," *Computational Statistics & Data Analysis*, 51, 2542–2558.

Kneip, A. and Ramsay, J. (2008), "Combining registration and fitting for functional models," *Journal of the American Statistical Association*, 103, 1155–1165.

Krivobokova, T., Crainiceanu, C., and Kauermann, G. (2008), "Fast adaptive penalized splines," *Journal of Computational and Graphical Statistics*, 17, 1–20.

Lang, S. and Brezger, A. (2004), "Bayesian P-splines," *Journal of Computational and Graphical Statistics*, 13, 183–212.

Ramsay, J. and Li, X. (2002), "Curve registration," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60, 351–363.

Ramsay, J. and Silverman, B. (2005), *Functional Data Analysis*, Springer.

Ritman, E. (2004), "Micro-computed tomography-current status and developments," *Annu. Rev. Biomed. Eng.*, 6, 185–208.

Rossi, M., Casali, F., Romani, D., Bondioli, L., Macchiarelli, R., and Rook, L. (2004), "MicroCT scan in paleobiology: Application to the study of dental tissues," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 213, 747–750.

Ruppert, D. and Carroll, R. (1999), "Penalized regression splines," Tech. rep., Cornell University Operations Research and Industrial Engineering.

— (2000), "Theory & Methods: Spatially-adaptive penalties for spline fitting," *Australian & New Zealand Journal of Statistics*, 42, 205–223.

Ruppert, D., Wand, M., and Carroll, R. (2003), *Semiparametric Regression*, Cambridge University Press.

Spiegelhalter, D., Best, N., Carlin, B., and Van Der Linde, A. (2002), "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639.

Staudenmayer, J., Ruppert, D., and Buonaccorsi, J. (2008), "Density estimation in the presence of heteroscedastic measurement error," *Journal of the American Statistical Association*, 103, 726–736.

Telesca, D. and Inoue, L. (2008), "Bayesian hierarchical curve registration," *Journal of the American Statistical Association*, 103, 328–339.

Thali, M., Taubenreuther, U., Karolczak, M., Braun, M., Brueschweiler, W., Kalender, W., Dirnhofer, R., et al. (2003), "Forensic microradiology: Micro-computed tomography (Micro-CT) and analysis of patterned injuries inside of bone," *Journal of Forensic Sciences*, 48, 1336–1342.

von Wittenau, A., Logan, C., Aufderheide III, M., and Slone, D. (2002), "Blurring artifacts in megavoltage radiography with a flat-panel imaging system: Comparison of Monte Carlo simulations with measurements," *Medical Physics*, 29, 2559.

The final version of this article appeared in the journal *Technometrics*:
http://amstat.tandfonline.com/loi/tech#.VlO_habPe3I

Full citation:
Lennox, Kristin P., and Lee G. Glascoe. "A Bayesian Measurement Error Model for Misaligned Radiographic Data." *Technometrics* 55.4 (2013): 450-460.